# Supporting Information

Arndt Schimmelmann[1]*, Haiping Qi[2], Tyler B. Coplen[2], Willi A. Brand[3], Jon Fong[1], Wolfram Meier-Augenstein[4], Helen F. Kemp[4], Blaza Toman[5], Annika Ackermann[6], Sergey Assonov[7], Anita Aerts-Bijma[8], Ramona Brejcha[9], Yoshito Chikaraishi[10], Tamim Darwish[11], Martin Elsner[9], Matthias Gehre[12], Heike Geilmann[3], Manfred Gröning[7], Jean-François Hélie[13], Sara Herrero-Martín[12], Harro A.J. Meijer[8], Peter E. Sauer[1], Alex L. Sessions[14], Roland A. Werner[6]

[1] Department of Geological Sciences, Indiana University, 1001 E 10th Street, Bloomington, Indiana 47405, USA. Fax (812) 855-7899 (A. Schimmelmann). aschimme@indiana.edu; jfong@indiana.edu; pesauer@indiana.edu

[2] U.S. Geological Survey, 431 National Center, Reston, Virginia 20192, USA. haipingq@usgs.gov; tbcoplen@usgs.gov

[3] Max-Planck-Institute for Biogeochemistry, Beutenberg Campus, P.O. Box 100164, 07701 Jena, Germany. wbrand@bgc-jena.mpg.de; geilmann@bgc-jena.mpg.de

[4] Stable Isotope Forensics Lab, James Hutton Institute, Invergowrie, Dundee, DD2 5DA, United Kingdom. w.meier-augenstein@rgu.ac.uk; helen.felicity.kemp@gmail.com
Present address: School of Pharmacy & Life Sciences, Robert Gordon University, The Sir Ian Wood Building, Aberdeen, AB10 7GJ, United Kingdom. w.meier-augenstein@rgu.ac.uk

[5] National Institute of Standards and Technology (NIST) Information Technology Laboratory, 100 Bureau Drive, M/S 8980, Gaithersburg, Maryland 20899-8980, USA. blaza.toman@nist.gov

[6] Institut für Agrarwissenschaften, ETH Zürich, LFW C 48.1, Universitätsstrasse 2, 8092 Zürich, Switzerland. annika.ackermann@usys.ethz.ch; roland.werner@usys.ethz.ch

[7] Terrestrial Environment, Laboratory, Environmental Laboratories, Department of Nuclear Applications, International Atomic Energy Agency (IAEA), 1400 Vienna, Austria. S.Assonov@iaea.org; M.Groening@iaea.org

[8] Centre for Isotope Research (CIO), Energy and Sustainability Research Institute Groningen (ESRIG), University of Groningen, Nijenborgh 4, 9747 AG Groningen, Netherlands. a.t.aerts-bijma@rug.nl; h.a.j.meijer@rug.nl

[9] Helmholtz Zentrum München, Institut für Grundwasserökologie, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. martin.elsner@helmholtz-muenchen.de; ramona.brejcha@helmholtz-muenchen.de

[10] Department of Biogeochemistry, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-Cho, Yokosuka, 237-0061, Japan. ychikaraishi@jamstec.go.jp

[11] National Deuteration Facility, Australian Nuclear Science and Technology Organisation, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia. tde@ansto.gov.au

[12] Department for Isotope Biogeochemistry, Helmholtz-Centre for Environmental Research (UFZ), Permoserstrasse 15, 04318 Leipzig, Germany. matthias.gehre@ufz.de; sara.herrero@zalf.de

[13] Centre de recherche GEOTOP, Département des sciences de la Terre et de l'atmosphère, Université du Québec à Montréal, Canada. helie.jean-francois@uqam.ca

[14] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. als@gps.caltech.edu

# New organic reference materials for hydrogen, carbon, and nitrogen stable isotope-ratio measurements: caffeines, *n*-alkanes, fatty acid methyl esters, glycines, L-valines, polyethylenes, and oils

**Abstract:**

An international project developed, quality-tested, and determined isotope–$\delta$ values of 19 new organic reference materials (RMs) for hydrogen, carbon, and nitrogen stable isotope-ratio measurements, in addition to analyzing pre-existing RMs NBS 22 (oil), IAEA-CH-7 (polyethylene foil), and IAEA-600 (caffeine). These new RMs enable users to normalize measurements of samples to isotope–$\delta$ scales. The RMs span a range of $\delta^{2}H_{VSMOW\text{-}SLAP}$ values from –210.8 to +397.0 mUr or ‰, for $\delta^{13}C_{VPDB\text{-}LSVEC}$ from –40.81 to +0.49 mUr, and for $\delta^{15}N_{Air}$ from –5.21 to +61.53 mUr. Many of the new RMs are amenable to gas and liquid chromatography. The RMs include triads of isotopically contrasting caffeines, $C_{16}$ *n*-alkanes, *n*-$C_{20}$-fatty acid methyl esters (FAMEs), glycines, and L-valines, together with polyethylene powder and string, one *n*-$C_{17}$-FAME, a vacuum oil (NBS 22a) to replace NBS 22 oil, and a $^{2}$H-enriched vacuum oil. Eleven laboratories from 7 countries used multiple analytical approaches and instrumentation for 2-point isotopic calibrations against international primary measurement standards. The use of reference waters in silver tubes allowed direct calibration of $\delta^{2}$H values of organic materials against isotopic reference waters following the principle of identical treatment. Bayesian statistical analysis yielded the mean values reported here. New RMs are numbered from USGS61 through USGS78, in addition to NBS 22a. Due to exchangeable hydrogen, amino acid RMs currently are recommended only for carbon- and nitrogen-isotope measurements. Some amino acids contain $^{13}$C and carbon-bound organic $^{2}$H-enrichments at different molecular sites to provide RMs for potential site-specific isotopic analysis in future studies.

## Overall contents of the Supplemental Information:

(i) A separate Microsoft Excel® file "Schimmelmann_et_al_color.xlsx" with 74 Tables (i.e. sheets), S1 through S74, is offering a comprehensive listing of all data with detailed graphic overviews; Table S1 (i.e. the first sheet of the Excel® file) serves as an explanatory "Contents" page for the Excel® file.

(ii) The same separate Microsoft Excel® file without color "Schimmelmann_et_al_bw.xlsx" is offered for printing.

(iii) This Microsoft Word file contains the following 6 components:

(1) Table S75 offering details in addition to Table 1 of the main publication: For each material and isotope system, the number *m* of reporting laboratories and the total number *n* of measurements are indicated.

(2) Method to incorporate uncertainty of measurement results and uncertainty of each substance used for normalization to calculate a combined uncertainty for each of the 22 materials and 11 laboratories in this study.

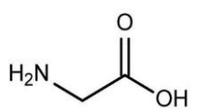
(3) Bayesian statistical analysis of stable isotope data to arrive at mean values.

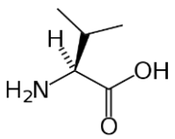
(4) Statistical model for combining $\delta^{13}$C data from regular (NBS 22a) and $^{2}$H-enriched (USGS78) vacuum oils in order to arrive at identical $\delta^{13}$C values for both oils.

(5) Comparison of a user's measurement with the RM $\delta$ value.

(6) Statistical model to calculate combined standard uncertainties for property values of the new RMs including heterogeneity effects and calibration uncertainty.

**(1) Table S75:** Alphabetical listing of new and pre-existing (*) organic reference materials with Bayesian Random Effects mean values with associated combined standard uncertainties.
[The unit mUr is synonymous with deprecated ‰; *m*, number of reporting laboratories; *n*, total number of measurements; n.a., not applicable; FAME, fatty acid methyl ester. The $\delta^2H_{VSMOW\text{-}SLAP}$ values are normalized to a scale in which $\delta^2H$ of SLAP = –428 mUr; the $\delta^{13}C$ values are normalized to a scale in which NBS 19 is +1.95 mUr and LSVEC is –46.6 mUr; the $\delta^{15}N$ values are normalized to a scale in which the $\delta^{15}N$ of atmospheric nitrogen = 0 mUr and that of USGS32 = +180 mUr.]

| Reference ID | Chemical name | Structure or composition of material | Ring-test reference values with combined standard uncertainties (mUr or ‰) | | |
|---|---|---|---|---|---|
| | | | $\delta^2H_{VSMOW\text{-}SLAP}$ | $\delta^{13}C_{VPDB\text{-}LSVEC}$ | $\delta^{15}N_{Air}$ |
| USGS61 | caffeine |  | +96.9 ± 0.9<br>*m=5; n=53* | –35.05 ± 0.04<br>*m=9; n=114* | –2.87 ± 0.04<br>*m=8; n=93* |
| USGS62 | caffeine | | –156.1 ± 2.1<br>*m=5; n=64* | –14.79 ± 0.04<br>*m=9; n=105* | +20.17 ± 0.06<br>*m=9; n=96* |
| USGS63 | caffeine | | +174.5 ± 0.9<br>*m=5; n=55* | –1.17 ± 0.04<br>*m=9; n=103* | +37.83 ± 0.06<br>*m=9; n=99* |
| IAEA-600* | caffeine | | –156.1 ± 1.3<br>*m=5; n=54* | –27.73 ± 0.04<br>*m=7; n=59* | +1.02 ± 0.05<br>*m=7; n=60* |
| USGS64 | glycine |  | *no values indicated due to the presence of exchangeable hydrogen* | –40.81 ± 0.04<br>*m=9; n=89* | +1.76 ± 0.06<br>*m=9; n=98* |
| USGS65 | glycine | | | –20.29 ± 0.04<br>*m=8; n=86* | +20.68 ± 0.06<br>*m=9; n=92* |
| USGS66 | glycine | | | –0.67 ± 0.04<br>*m=9; n=96* | +40.83 ± 0.06<br>*m=9; n=92* |
| USGS67 | *n*-hexadecane | $C_{16}H_{34}$ | –166.2 ± 1.0<br>*m=9; n=163* | –34.50 ± 0.05<br>*m=8; n=99* | n.a. |
| USGS68 | *n*-hexadecane | | –10.2 ± 0.9<br>*m=9; n=147* | –10.55 ± 0.04<br>*m=8; n=91* | n.a. |
| USGS69 | *n*-hexadecane | | +381.4 ± 3.5<br>*m=9; n=132* | -0.57 ± 0.04<br>*m=7; n=86* | n.a. |

| | | | | | |
|---|---|---|---|---|---|
| USGS70 | icosanoic acid methyl ester ($C_{20}$ FAME) | $C_{20}H_{39}OOCH_3$ | –183.9 ± 1.4<br>*m=9; n=116* | –30.53 ± 0.04<br>*m=9; n=77* | n.a. |
| USGS71 | icosanoic acid methyl ester ($C_{20}$ FAME) | | –4.9 ± 1.0<br>*m=9; n=118* | –10.50 ± 0.03<br>*m=9; n=65* | n.a. |
| USGS72 | icosanoic acid methyl ester ($C_{20}$ FAME) | | +348.3 ± 1.5<br>*m=9; n=130* | –1.54 ± 0.03<br>*m=8; n=62* | n.a. |
| USGS73 | L-valine |  | *no values indicated due to the presence of exchangeable hydrogen* | –24.03 ± 0.04<br>*m=9; n=130* | –5.21 ± 0.05<br>*m=9; n=91* |
| USGS74 | L-valine | | | –9.30 ± 0.04<br>*m=7; n=94* | +30.19 ± 0.07<br>*m=7; n=68* |
| USGS75 | L-valine | | | +0.49 ± 0.07<br>*m=5; n=23* | +61.53 ± 0.14<br>*m=6; n=29* |
| USGS76 | methylheptadecanoate ($C_{17}$ FAME) | $C_{17}H_{33}OOCH_3$ | –210.8 ± 0.9<br>*m=9; n=131* | –31.36 ± 0.04<br>*m=8; n=93* | n.a. |
| IAEA-CH-7* | polyethylene foil | $(C_2H_4)_n$ | –99.2 ± 1.2<br>*m=8; n=143* | –32.14 ± 0.05<br>*m=5; n=49* | n.a. |
| USGS77 | polyethylene powder (also extruded string) | $(C_2H_4)_n$ | –75.9 ± 0.6<br>*m=8; n=199* | –30.71 ± 0.04<br>*m=8; n=81* | n.a. |
| NBS 22* | oil | n.a. | –117.2 ± 0.6<br>*m=8; n=186* | –30.02 ± 0.04<br>*m=8; n=72* | n.a. |
| NBS 22a | vacuum oil, regular | n.a. | –120.4 ± 1.0<br>*m=8; n=203* | –29.72 ± 0.04<br>*m=8; n=103* | n.a. |
| USGS78 | vacuum oil, $^2$H-enriched | n.a. | +397.0 ± 2.2<br>*m=7; n=200* | –29.72 ± 0.04<br>*m=7; n=80* | n.a. |

## (2) Method to incorporate uncertainty of measurement results and uncertainty of each substance used for normalization to calculate a combined uncertainty for each of the 22 materials and 11 laboratories in this study

(by Tyler B. Coplen, U.S. Geological Survey, 431 National Center, Reston, Virginia, January 29, 2016)

There are several sources of uncertainty that should be considered in the uncertainty budget to determine a combined uncertainty for each laboratory's measurements of each of the 22 materials analyzed in this study. First, the measurement uncertainty of each of the 22 reference materials (RMs) reported by laboratories is found in column C of each of the spreadsheets labeled "S3 $\delta^2$H RM uncertainties", "S4 $\delta^{13}$C RM uncertainties", and "S5 $\delta^{15}$N RM uncertainties", found in the Microsoft Excel® file in Supporting Information. The measurement uncertainties are the standard deviations calculated from individual measurement results of laboratories listed on spreadsheets S63 through S73.

A second source of uncertainty arises from the fact that the RMs used for normalization of results of the 22 materials typically have an uncertainty greater than zero, and the uncertainty of each of the RMs used for normalization should be considered. For internationally distributed RMs, the uncertainty values come from Brand et al. [1]. Consider the $\delta^2$H normalization of USGS61 caffeine by Indiana University (IU; see the worksheet labeled "S3 $\delta^2$H RM uncertainties", in the Excel® file in Supporting Information). Indiana University used four reference materials VSMOW2, SLAP2, +799.9 $H_2O$, and 'other2' (row 3), having $\delta^2H_{VSMOW\text{-}SLAP}$ values, respectively, of 0, –427.5, +799.9, and +505.5 mUr (row 2), and having uncertainties, respectively, of 0.4, 0.4, 0.5, and 1.7 mUr (row 4). The measured mean $\delta^2H_{VSMOW\text{-}SLAP}$ value of USGS61 is +98.0 mUr (cell B6). The absolute value of the difference between +98.0 mUr and the $\delta$ values of each of the references is 701.9, 407.5, 505.5, and 98 mUr. Therefore, the relative weighted fraction of each of these is 0.089, 0.153, 0.123, and 0.635, which sum to one. The estimated uncertainty is the sum of the product of each of these fractions and its respective uncertainty: $0.089\times0.5 + 0.153\times1.7 + 0.123\times0.4 + 0.635\times0.4 = 0.607$ mUr, which rounds to 0.6 mUr, the value found in cell Y6.

A third source of uncertainty is the measurement uncertainty of the RMs used for normalization. A complication arises in that most laboratories did not provide individual measurement results of RMs used for normalization, and some laboratories provided no information on measurement uncertainty of RMs used for normalization. Therefore, we need to develop an algorithm to estimate this source of uncertainty. We assume that this measurement uncertainty ought to be related to the measurement uncertainty of the 22 materials analyzed in this study. After reviewing the mean standard deviations of RMs used for normalization provided by 5 laboratories (CH, spreadsheet S64; Leipzig, spreadsheet S68; Munich, spreadsheet S69; NL, spreadsheet S70; UK, spreadsheet S74), we conclude that a satisfactory estimate for the measurement uncertainty of RMs used for normalization is the median of the measurement uncertainties of the materials analyzed in this study. For example, consider the measurement uncertainties of the nine materials whose $\delta^{15}N_{N2\text{-}Air}$ values were determined by USGS (Table S76). The standard deviations of these values range from 0.037 to 0.132 mUr. The median $\delta^{15}N_{N2\text{-}Air}$ value of 0.071 mUr is selected for the measurement uncertainty of RMs used for nitrogen-isotope normalization. Thus, the value of 0.071 appears in cell P15 of the spreadsheet "S5 $\delta^{15}$N RM uncertainties" of the Excel® file in Supporting Information. Table S77 lists the measurement uncertainty of each of the reference materials for each of the laboratories used for normalization of the materials in this study.

**Table S76:** Measurement uncertainty of $\delta^{15}N$ analyses of materials analyzed by USGS in this study

[Median value of 0.071 mUr, with gray background, is selected for the estimated $\delta^{15}N$ measurement uncertainty for all RMs used for $\delta^{15}N$ normalization of the 9 materials analyzed by USGS.]

| Name | $\delta^{15}N_{N2\text{-}Air}$ (mUr) | Std. Dev. (mUr) |
|---|---|---|
| USGS65 | 20.731 | 0.037 |
| USGS61 | -2.829 | 0.046 |
| USGS64 | 1.780 | 0.048 |
| IAEA-600 | 0.975 | 0.055 |
| USGS63 | 37.974 | 0.071 |
| USGS62 | 20.213 | 0.109 |
| USGS66 | 40.989 | 0.109 |
| USGS74 | 30.210 | 0.116 |
| USGS73 | -5.234 | 0.132 |

**Table S77:** Estimated measurement uncertainty of reference materials used for normalization of the 22 materials in this study

| | Estimated Measurement Uncertainty | | |
|---|---|---|---|
| Laboratory | $\delta^{2}H$ (mUr) | $\delta^{13}C$ (mUr) | $\delta^{15}N$ (mUr) |
| Caltech | 2.4 | | |
| CH | 1.2 | | |
| IU | 2.0 | 0.021 | 0.030 |
| Jena | 1.4 | 0.064 | 0.120 |
| JP | | 0.049 | 0.046 |
| Leipzig | 1.4 | 0.087 | 0.052 |
| Munich | 1.9 | 0.026 | 0.041 |
| NL | 1.9 | 0.033 | 0.061 |
| UK | 0.8 | 0.051 | 0.072 |
| UQAM | | 0.105 | 0.248 |
| USGS | 0.9 | 0.072 | 0.071 |

The combined uncertainty is the square root of the square of the three uncertainties discussed above. For example, for the Indiana University $\delta^{2}H_{VSMOW\text{-}SLAP}$ measurement of USGS61 caffeine, the measurement uncertainty (cell C6), the uncertainty of each of the RMs used for normalization (cell Y6), and the estimated measurement uncertainty of RMs used for normalization (cell Z6) are 1.5, 0.6, and 2.0 mUr,

respectively. The combined uncertainty (cell AA6) = $(1.5^2 + 0.6^2 + 2.0^2)^{0.5}$ = 2.6 mUr. The combined uncertainties of $\delta^{13}C_{VPDB\text{-}LSVEC}$ and $\delta^{15}N_{Air\text{-}N2}$ measurements are determined in an equivalent manner.

**Reference:**

[1] Brand WA, Coplen TB, Vogl J, Rosner M, Prohaska T (2014) Assessment of international reference materials for stable isotope ratio analysis. *Pure and Applied Chemistry* **86** (3), 425 – 467. http://dx.doi.org/10.1515/pac-2013-1023

## (3) Bayesian statistical analysis of stable isotope data to arrive at mean values

(by Blaza Toman, National Institute of Standards and Technology (NIST), Information Technology Laboratory, Gaithersburg, Maryland, November 16, 2015)

### Data

Only 2-point-calibrated data were accepted. All $\delta^2$H values were accepted except for those that had been measured on-line with high-temperature-conversion methodologies (e.g., conventional TC/EA) for nitrogen-containing compounds in the absence of chromium to scavenge nitrogen. All $\delta^{13}$C values were accepted except for the data sets of four materials from a single laboratory that had used peak jumping when measuring $\delta^{13}$C and $\delta^{15}$N values in the same runs. The group voted to eliminate these four sets of $\delta^{13}$C values as outliers. The group also voted to eliminate the $\delta^{15}$N values of one material from a single laboratory due to outlier status. All accepted data, or sets of observations, were used for each estimate (i.e. mean value).

There is one exception to the statistical treatment; the $\delta^{13}$C values of the regular vacuum oil NBS 22a and the $^2$H-enriched vacuum oil USGS78 were harmonized, as described in section 2 in Supplementary Information.

### Statistical Model

For each compound and isotope system there were delta values $Y_i$, with accompanying combined uncertainties $u(Y_i)$,computed as described in section (2), and sample sizes $n_i$, for each laboratory *i* that made the measurements. The values $Y_i$ were arithmetic means of the $n_i$ measurements. The common delta value for each compound and isotope system, $\mu_{compound}$, was estimated using the following statistical model:

1. For each laboratory, the observations $Y_i$ were assumed to be independent Gaussian random variables with mean $m_i$ and variance $\frac{u(Y_i)^2}{n_i}$.
2. The means $m_i$ were themselves Gaussian random variables with mean $\mu_{compound}$ and variance $\sigma^2_{compound}$.

Assumption 1 considered the uncertainties $u(Y_i)$ to be due to random measurement errors and so be of Type A [1] and subject to division by $n_i$. Assumption 2 quantified, in the form of $\sigma_{compound}$, additional uncertainty evident in the data as between laboratory variability (see [2, 3] for further discussion of such random effects models). The estimation of $\mu_{compound}$ was carried out using Bayesian methods (Markov

Chain Monte Carlo [4]), coded in OpenBUGS [5]. Bayesian models require additional inputs called prior distributions for the parameters $\mu_{compound}$, and $\sigma_{compound}$. In the absence of any additional information, as was the case here, these distributions are chosen to be non-informative, that is, chosen to not affect the results of the analysis.

The fit of the model was examined using posterior predictive probabilities $P(Y_{ip} > Y_i)$, where $Y_{ip}$ are predicted values according to the statistical model above. These probabilities, usually called Bayesian posterior predictive *p*-values [6], measure how likely it is to obtain the value $Y_i$ given the statistical model and all of the observed data. Predictive p-values for data points that fit the model well should be around 0.5 [6], that is, about half the time the predicted value should be larger than the observed and half the time smaller. In this analysis, all observations satisfied this criterion.

The common delta values were estimated by the posterior means of $\mu_{compound}$. Because it could be argued that $u(Y_i)$ is not strictly due to measurement errors, the standard and expanded uncertainties were estimated by the posterior standard deviation and by the half width of the 95% Highest Probability Density interval for $\mu_{compound}$ using a statistical model where the variance in Assumption 1 was $u(Y_i)^2$ instead of $\frac{u(Y_i)^2}{n_i}$. This produced conservative standard and expanded uncertainties.

**References:**


[1] JCGM 100 (2008) Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement (Sevres, France: BIPM).

[2] Toman B (2007) Bayesian approaches to calculating a reference value in key comparison experiments. *Technometrics* **49** (1), 81 – 87. http://dx.doi.org/10.1198/004017006000000273

[3] Toman B, Fischer J, Elster C (2012) Alternative analyses of measurements of the Planck constant. Metrologia 49, 567 – 571.

[4] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis*. Chapman & Hall, Boca Raton, 3rd edition.

[5] Lunn D, Spiegelhalter D, Thomas A, Best N (2009) The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine* **28**, 3049 – 3082. http://dx.doi.org/10.1002/sim.3680

[6] Gelman A, Meng XL, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6**, 733 – 807. http://www.stat.columbia.edu/~gelman/research/published/A6n41.pdf


## (4) Statistical model for combining $\delta^{13}C$ data from regular (NBS 22a) and $^2H$-enriched (USGS78) vacuum oils in order to arrive at identical $\delta^{13}C$ values for both oils

(by Blaza Toman, National Institute of Standards and Technology (NIST), Information Technology Laboratory, Gaithersburg, Maryland, September 11, 2015)

*Rationale:* The carbon isotopic composition of both vacuum oils has to be essentially identical because only trace amounts (250 mg) of perdeuterated *n*-tetracosane were added to 3.35 kg of regular vacuum oil (NBS 22a) to generate the $^2H$-enriched vacuum oil (USGS78). Both oils were measured separately.

The difference between resulting average $\delta^{13}$C values (NBS 22a: –29.73 ± 0.04 mUr; USGS78: –29.71 ± 0.04 mUr) was close to the precision of the respective measurements. It was decided to statistically pool all $\delta^{13}$C measurements of both oils to arrive at an overall mean value, but it was necessary to take into account that the regular oil had been measured with better precision and more often (n=103) than the $^{2}$H-enriched vacuum oil (n=80). The following text describes the statistical procedure.

Let $\mu_{regoil}$ be the measurand for the regular vacuum oil and $\mu_{enoil}$ be the measurand for the $^{2}$H-enriched vacuum oil. Further, let $\mu_{consensus}$ be a mean value for vacuum oil. The following statistical model was used to combine the results of the regular and enriched oils:

1. The measurements of regular oil $Y_{1i}$ were assumed to be independent Gaussian random variables with mean $m_{1i}$ and variance $\frac{u(Y_{1i})^2}{n_{1i}}$, for laboratories $i = 1,\ldots, N_{regoil}$.

   The measurements of enriched oil $Y_{2i}$ were assumed to be independent Gaussian random variables with mean $m_{2i}$ and variance $\frac{u(Y_{2i})^2}{n_{2i}}$, for laboratories $i = 1,\ldots, N_{enoil}$.
2. The means $m_{1i}$ were Gaussian random variables with mean $\mu_{regoil}$ and variance $\sigma^2_{regoil}$, means $m_{2i}$ were Gaussian random variables with mean $\mu_{enoil}$ and variance $\sigma^2_{enoil}$.
3. Further, $\mu_{regoil}$ and $\mu_{enoil}$ were Gaussian random variables with mean $\mu_{consensus}$ and variance $\sigma^2_{consensus}$.

This model accounted for laboratory measurement uncertainty using $u(Y_{1i})$ and $u(Y_{2i})$, for additional between laboratory uncertainty using $\sigma_{regoil}$ and $\sigma_{enoil}$, and for uncertainty due to variability between oils using $\sigma_{consensus}$.

The resulting estimated $\delta^{13}$C value for both oils is: –29.723 mUr, standard uncertainty: 0.038 mUr, and expanded uncertainty 0.076 mUr.

## (5) Comparison of a user's measurement with the RM $\delta$-value.

(by Blaza Toman, National Institute of Standards and Technology (NIST), Information Technology Laboratory, Gaithersburg, Maryland, February 18, 2016)

Suppose that an RM $\delta$-value in Table 2 is given as $X$, and the combined standard uncertainty as $u(x)$. Suppose that a user laboratory performs one measurement of this material and obtains a value $Y$. To compare their result to the RM $\delta$-value they would compute the difference $d = Y - X$ and find the accompanying measurement uncertainty as $u(d) = \sqrt{u(x)^2 + u(y)^2}$. Then the user's result is in good agreement with the RM $\delta$-value when $|d| \le 2u(d)$.

Clearly, to go through with this process the user laboratory needs a value for $u(y)$. For this reason, it is better to obtain several replicated measurements, $Y_1, Y_2, \ldots, Y_n$, compute the average $\bar{Y}$, and the standard

deviation $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}$. Then $d = \bar{Y} - X$, $u(d) = \sqrt{u(x)^2 + \frac{s^2}{n}}$, and one would check whether $|d| \le 2u(d)$. When replicate measurements are not possible, the user laboratory can estimate their measurement uncertainty using a standard deviation of their past replicate measurements, or using other means such as expert knowledge. An alternative is an approximation of $u(y)$ based on the fact that for the laboratories that participated in this study, the average values of $u(y)$ were between 2 to 3 times $u(x)$. If the approximation that $u(y) = 3u(x)$ was used then the user's result would be in good agreement with the RM $\delta$-value when $|d| \le 6.2u(x)$.

## (6) Statistical model to calculate combined standard uncertainties for property values of the new reference materials including heterogeneity effects and calibration uncertainty.

(by Sergey Assonov and Manfred Gröning, International Atomic Energy Agency (IAEA), Terrestrial Environment Laboratory, Seibersdorf, Austria, 2 March 2016)

*Rationale:* As calculated uncertainties for newly assessed materials in Table S75 herein and Table 2 of the publication are sometimes lower than the uncertainty assigned to the existing reference materials in use for the calibration, clarification is deemed necessary on the procedure to calculate combined standard uncertainties for delta-values of reference materials. Here we outline an alternative approach taking into account uncertainty components in a different manner that would result in larger combined standard uncertainties.

The calibration of new reference materials (RMs) was performed by using data from eleven experienced laboratories following individual best practices. Data were evaluated to provide best estimates for the isotopic reference values as listed in Table S75 herein and Table 2 of the publication.

The study and the data evaluation addressed uncertainty components in the following way (for exact details see Supporting Information components 2 and 3):

1. Laboratory analytical reproducibility was calculated for each property value ($\delta^2$H, $\delta^{13}$C, $\delta^{15}$N) of each new RM considering the whole data set for each lab (data on normalized 2-point calibration) and the calculation of mean and standard deviation;
2. The uncertainty of the 2-point calibration approach was realized as based on RMs used for calibration by two terms, namely the assigned calibration RM uncertainty $u_{cal}$ (as stated in its reference document) and the analytical uncertainty of measurements performed on these RMs used for calibration.
3. The three components above were combined, thus producing for each material and for each laboratory the mean value (based on $N$ aliquots analyzed in each lab) and its uncertainty $u_i$ (1-σ level).
4. The data sets from all $n$ participating labs, for each compound (data provided as described above) and including a term estimating Between-Laboratory uncertainty, were combined by Bayesian statistics. The values produced by the data treatment are considered as the best estimation of the property values as provided in Tables 2 and S75.

This provides the information to calculate the combined standard uncertainty $U$ for each material by combining all $u_i^2$ values (step 3 above) of individual laboratories $i$, which each one is containing two uncertainty contributions from calibration materials according to point 2 above, and the Between-Laboratory uncertainty.

The combined standard uncertainty $U$ according to Table 2 for each new RM is sufficient to demonstrate the compatibility of laboratory performance, their mutual consistency, and to show the high quality of analytical work and of materials produced.

However, an alternative approach as presented here additionally takes into account the following two aspects:

1) All laboratories calibrated their measurements by use of similar calibration standards. All laboratories have this common uncertainty component $u_{cal}$, which can be considered as a nearly perfect covariance and therefore needs to be dealt with differently as in common uncertainty propagation. Especially it means that the common uncertainty component $u_{cal}$ may not be subject to division by the square root of $n$, as is implicitly done in the numerical procedure in step 4 above. This is necessary here due to the covariance and important for the purpose of calibration of reference materials, and it would not be important in case of a comparison of laboratory results. It can be understood easiest if one considers a hypothetical case of the same calibration materials used by each laboratory ($u_{cal,i} = u_{cal}$ being equal for all laboratories $i$). This creates a perfect covariance in that uncertainty component $u_{cal}$ is common for all laboratories. This uncertainty component $u_{cal}$ is completely independent from the number of participating laboratories, and therefore should not be diminished during the uncertainty propagation [R1, R2]. This can be ensured by not including the uncertainty component $u_{cal}$ related to the assigned calibration uncertainty in the evaluation of all other individual laboratory uncertainty components $u_i$. After all other uncertainty components are evaluated following steps 1–4 above producing an uncertainty $u_{meas}$, the calibration uncertainty $u_{cal}$ component can be added afterwards as (Eq. 1).

$$U = \sqrt{u_{\mathrm{meas}}^2 + u_{\mathrm{cal}}^2} \qquad (1)$$

This will ensure that the full magnitude of this uncertainty component is maintained in $U$.

1) The second aspect is the necessary heterogeneity test for each new reference material after bottling into individual units. This uncertainty component ensures that possible deviations of the isotopic composition in individual bottles are included. This uncertainty component would be mandatory to be compiled and included in the uncertainty budget towards compliance with ISO Guide 35 [R3] applying GUM principles [R4] on reference materials. Often this uncertainty contribution is not negligible and may contribute significantly to the final uncertainty [R5, R6, R7]. This assessment was not possible to achieve within the recent calibration scheme performed in this study. Adherence to ISO Guide 35 [R3] is not required for this scientific publication in *Analytical Chemistry*. A possible inclusion of the heterogeneity contribution $u_{het}$ to $U$ follows the case of $u_{cal}$ above:

$$U = \sqrt{u_{\mathrm{meas}}^2 + u_{\mathrm{cal}}^2 + u_{\mathrm{het}}^2} \qquad (2)$$

Equation 2 provides then the final standard combined uncertainty $U$ with all terms properly included, to be assigned to the property value of the reference material.

The approach as explained under 1) with a common $u_{cal}$ is strictly applicable only in the case that all laboratories use the same calibration materials [R2, p. 468]. However, due to the close relationship of the existing calibration materials used by the participating laboratories (due to former RM cross-calibration and in view of the similarity of RM uncertainty values), this approach is justified.

A special case is given by the recently discovered drift in $\delta^{13}C$ property values of the two reference materials LSVEC and USGS41 used as second RM for normalization by several laboratories. Their uncertainty values have subsequently been expanded to enclose the possible bias of individual units. As they were used by a significant number of the laboratories, that influence is visible on final combined uncertainties calculated according to Eq. 2, using the same median approach to estimate calibration uncertainties as performed in the original evaluation (in component 2 of this Supporting Information).

The second uncertainty term $u_{cal}$ in Eq. 1 is negligible in case of $\delta^{2}H$, as it is in the order of 0.3 mUr only, compared to measurement reproducibility much above 1 mUr for most samples. However, this $u_{cal}$ term is highly relevant for $\delta^{13}C$ and $\delta^{15}N$ though.

Applying the same model as in component 2 to approach the calibration/normalization uncertainty by the median value of $u_{cal,i}$ uncertainties, according to Eq. 2 it results in significantly larger combined standard uncertainties *U*.

Two examples are given below for the re-calculation of the combined standard uncertainty *U* according to Eq. 1 (not including any heterogeneity component $u_{het}$):

a) NBS22a – for $\delta^{13}C$:
   Uncertainty *U* derived from evaluation scheme as for Table 2: ±0.04 mUr
   Uncertainty *U* derived by use of Eq. 1: **±0.10** mUr
b) USGS61 – for $\delta^{15}N$:
   Uncertainty derived from evaluation scheme as for Table 2: ±0.04 mUr
   Uncertainty derived by use of Eq. 1: **±0.08** mUr

These uncertainties may rise further in case of any isotopic heterogeneity between bottles $u_{het}$ would be detected by using Eq. 2.

It is therefore demonstrated that any possible heterogeneity of prepared individual aliquots for distribution can be assessed, and at the same time assessment of the calibration uncertainty can be determined by use of 2-point calibration with materials of lowest possible uncertainty. In this manner, these uncertainty data can be provided on reference sheets of isotopic reference materials.

**References:**


[R1] Pauwels, J., Lamberty, A., Schimmel, H., 1998. The determination of the uncertainty of reference materials certified by laboratory intercomparison. *Accreditation and Quality Assurance* **3**, 180-184. http://dx.doi.org/10.1007/s007690050218

[R2] van der Veen, A.M.H., Pauwels, J., 2000. Uncertainty calculations in the certification of reference materials. 1. Principles of analysis of variance. *Accreditation and Quality Assurance* **5**, 464-469. http://dx.doi.org/10.1007/978-3-662-05173-3_16

[R3] ISO, 2006. Reference materials -- General and statistical principles for certification (ISO Guide 35:2006), International Organization for Standardization, Geneva, Switzerland. 1-64.

[R4] JCGM, 2008. Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM), JCGM Joint Committee for Guides in Metrology, ISO, Geneva, Switzerland. 1-134.

[R5] Pauwels, J., Lamberty, A., Schimmel, H., 1998. Homogeneity testing of reference materials. *Accreditation and Quality Assurance* **3**, 51-55. http://dx.doi.org/10.1007/s007690050186

[R6] van der Veen, A.M.H., Linsinger, T., Pauwels, J., 2001. Uncertainty calculations in the certification of reference materials. 2. Homogeneity study. *Accreditation and Quality Assurance* **6**, 26-30. http://dx.doi.org/10.1007/978-3-662-05173-3_17

[R7] Linsinger, T.P.J., Pauwels, J., van der Veen, A.M.H., Schimmel, H., Lamberty, A., 2001. Homogeneity and stability of reference materials. *Accreditation and Quality Assurance* **6**, 20-25. http://dx.doi.org/10.1007/s007690000261